Cell, Volume *165*

Supplemental Information

The CDK-APC/C Oscillator Predominantly Entrains Periodic Cell-Cycle Transcription

Sahand Jamal Rahi, Kresti Pecani, Andrej Ondracka, Catherine Oikonomou, and Frederick R. Cross

Supplemental Experimental Procedures

Strains

Standard methods were used throughout. All strains are W303-congenic.

Name	cln1Δ	cln2Δ	cln2Δ:CLN2pr-Venus:TRP1	cln3Δ:LEU2	trp1Δ:TRP1:MET3-CLN2	clb1Δ-clb6Δ:KanMX	clb2Δ:GALL-CLB2:URA3-clb5Δ:KanMX	clb3Δ:TRP1	clb4Δ:his3:KanMX	swi5Δ:KanMX	HTB2-mCherry:HIS5	SIC1:SIC1pr-YFP:URA3	CLB2:CLB2pr-GFP:NatMX	SWI5-YFP:HIS3	CDC6:CDC6pr-YFP:NatMX	CYK3:CYK3pr-YFP:NatMX
2773-1D	X	X		X	X											
SJR17d4c	X	X		X	X	X	X	X	X		X					
SJR40a8c			X								X					
SJR14a4d	X		X	X	X						X					
SJR12a5a	X		X	X	X	X	X	X	X		X					
SJR78a13d	X		X		X	X	X	X	X		X					
SJR41a1a												X				
SJR5b	X	X		X	X							X				
SJR17a13c	X	X		X	X	X	X	X	X		X	X				
KP-A-4d	X	X		X	X	X	X	X	X	X	X	X				
SJR27a7b	X	X		X	X	X	X	X	X			X				
KP-Y-F4D											X		X			
SJR59-1	X	X		X	X								X			
SJR23-1	X	X		X	X	X	X	X	X		X		X			
SJR62a15a	X	X		X	X						X			X		
SJR47a3c	X	X		X	X	X	X	X	X		X			X		
SJR77a1	X	X		X	X						X				X	
SJR77a2	X	X		X	X	X	X	X	X		X				X	
SJR77a4	X	X		X	X	X	X	X	X	X	X				X	
SJR77b1	X	X		X	X						X					X
SJR77b2	X	X		X	X	X	X	X	X		X					X
SJR77b4	X	X		X	X	X	X	X	X	X	X					X

Name	cdc20Δ:MET3-HA3-CDC20:TRP1	ura3Δ:GAL1-CLB2kd:URA3	ADH1:ADH1pr-GAL4-rMR:URA3	swi5Δ:KanMX	CLN2:CLN2pr-GFP-CLN2PEST:HIS3	MYO1-mCherry:HIS3	SWI5-YFP:HIS3	HTB2-mCherry:HIS3	CLB2:CLB2pr-GFP-PEST:LEU2	SIC1:SIC1pr-YFP:URA3	CDC6:CDC6pr-YFP:NatMX	CYK3:CYK3pr-YFP:NatMX
BD143a-21c	X	X	X		X	X						
SJR63c2c	X				X	X						
ALG911	X	X	X			X	X	X				
SJR64a4c	X					X	X	X				
SJR44c12b	X	X							X			
SJR44a9d	X								X			
SJR19a11a	X	X	X							X		
KPA1-10b	X									X		
KPA1-2d	X	X		X						X		
KPE-A8D	X	X									X	
KPE-A8B	X		X								X	
KPE-B6C	X	X		X							X	
KPF-A4A	X	X										X
KPF-A7A	X		X									X
KPF-B3C	X	X		X								X

Name	trp1Δ:TRP1:GAL-SIC1(1X)	clb2Δ:LEU2	sic1Δ:HIS3	swi5Δ:KanMX	ace2Δ:URA3	APC-A (CDC16A:TRP1 CDC23A-HA CDC27A:kanR)	MYO1-GFP:KanMX	HTB2-mCherry:HIS5	ura3Δ:SIC1pr-YFP:URA3	clnΔ* CLB2-YFP:HIS3
SJR45a13c	X	X	X				X	X	X	
SJR45b12a	X	X					X	X	X	
SJR46a1b	X		X				X	X	X	
SJR45a13a	X						X	X	X	
SJR56a7c	X	X		X	X		X	X		
SJR55b1d	X	X		X			X	X		
SJR56a7a	X			X	X		X	X		
SJR55a3d	X			X			X	X		
SJR90c8d	X		X			X	X	X		
SJR90e5a	X					X		X		
SJR89f1a	X		X					X		X
SJR89d23b	X							X		X

We verify the CLN and CLB deletions by the absence of RNA-seq reads in our whole-transcriptome measurements. We check for all deletions, including CLN and CLB deletions, by PCR. The CLB deletions come from the Nasmyth lab. The SWI5-YFP gene comes from the Stillman lab [Sbia et al., 2008]. The SIC1pr-YFP, CDC6pr-YFP, and CYK3pr-YFP constructs come from the Colman-Lerner lab [Colman-Lerner et al., 2001].

Whole-transcriptome sequencing

clnΔ^* and clnΔ^*clbΔ^* cultures were grown in synthetic galactose medium without methionine (G-Met) overnight and subjected to the series of media changes that constitute the cyclin-depletion protocol. At time $0'$, each culture was split into two, one culture was left in synthetic glucose medium with methionine (D+Met, for Fig. 2 C) and MET-CLN2 was induced temporarily in the other by switching to synthetic glucose medium without methionine (D-Met) at $0'$ and adding methionine back at $90'$ (for Fig. 2 D). Samples were taken at $0'$, $30'$, $60'$, $120'$ from the culture that stayed in D+Met, and samples were taken at $0'$, $30'$, $60'$, $90'$, $120'$, $150'$, $180'$, and $240'$ from the culture, in which a switch to and from D-Met were performed at $0'$ and $90'$, respectively. (At $0'$ and $90'$, samples were taken right before changing media.) RNA was extracted using standard methods and cDNA libraries were prepared using the Illumina TruSeq RNA Sample Preparation Kit v2. Sequencing was carried out by Genewiz on an Illumina HiSeq2500 platform in 1x50 bp single-read configuration in Rapid Run mode. Reads were aligned to the Saccharomyces Cerevisiae S288C R64.1.1 reference genome ORFs (Engel et al, G3 (Bethesda) (2013)) using Bowtie 2 (Langmead and Salzberg, Nature Methods (2012)). FPKM scores were computed using eXpress (Roberts and Pachter, Nature Methods (2012)).

Genes not considered in transcriptome analysis in Fig. 2 C,D

130 cell-cycle regulated genes and their transcription factors were compiled from [Spellman et al., 1998, Colman-Lerner et al., 2001, Haase and Wittenberg, 2014]. The following genes were downregulated or not sufficiently upregulated during the MET-CLN2-induced cell cycle in clnΔ^* cells:

SBF cluster	FKS1, KAR4, KRE6, MNN1, TIP1
MBF cluster	RAD54, POL12, GIC2, DPB3, DBF4, MBP1
Histone cluster	SPT10
G2 cluster	CWP2, TIR1
CLB2 cluster	ASE1, DBF2, FAR1, SED1
ACE2 cluster	ASH1
SWI5 cluster	AGA1, CHS1, FUS1, MFA2, NIS1, PIR3, PST1, RAD3, SST2, STE2, TEC1
Mcm1 cluster	MCM1, MCM3, MCM4, MCM5, SWI4

Genes with potentially ambiguous regulation:

CLB2 cluster	MOB1 also has binding sites for Start TF Swi6 (Lee et al, Science (2002))
Swi5 cluster	PCL2 also has binding sites for Start TF Swi6 (Lee et al, Science (2002))

Genes irrelevant for the GTO model:

MAT cluster	STE12, FUS3, FAR1

Biological significance test

We consider fluctuations insignificant if they rise or drop by less than 10% of the wild-type amount. (We consider fluctuations in the range 11%-30% ambiguous.) We have an additional condition for the cases where in the CDK-APC/C block, gene activity is very high and declines slowly or is high and increases slowly. We set the treshold for the levels to return back to peak levels from above or to half-peak levels from below at 3 * cell cycle times = 200 min. Such a slow, steady change is considered 'biologically insignificant' in the sense that it does not resemble a cell cycle oscillation.

Single-cell/single gene measurements

We quantify the rise $\Lambda_\uparrow$ and drop $\Lambda_\downarrow$ of fluorescence by smoothing individual time courses (see below for smoothing parameters), and measuring the rise and drop in fluorescence as depicted in Fig. S3 A. The drop versus rise scatter plots are shown in the supplementary figures and the specific quantity (average rise $\bar{\Lambda}_\uparrow$ or average drop $\bar{\Lambda}_\downarrow$) that violates the threshold, if any, is stated there as well.

In the cdc20Δ experiments, CLN2pr-GFP returns to half its previous peak in 200', on average, and CLB2pr-GFP returns back to its previous peak in 320', on average, which we consider too slow to be reasonably called a cell-cycle-like oscillation (see Fig. S3 H,I). It is consistent with slow release from the MET-CDC20 cdc20Δ block since recovery always coincides with cell division and arrest.

Transcriptome measurements

It is possible that a fluctuation in one gene is so large that it is above 10% of the wild-type peak just by chance, or vice versa, below one of the thresholds (10% or 30%) just by chance. Many of the 91 cell cycle regulated genes that we consider are barely upregulated enough during the normal cell cycle (MET-CLN2-induced clnΔ^* cells) to pass the significance (p=0.01) threshold (see amplitude test for p values). Even if such a weakly upregulated gene is perfectly constant in a CDK-APC/C block, noise can easily put the constant background signal above 10% of the wild-type peak (for a gene that oscillates at p=0.01 compared to noise in the wild-type cell cycle, the probability that noise will produce a peak greater than 10% of the wild-type peak is 0.33). This problem could be solved by comparing the size of the fluctuation to the measurement noise but we wish to keep the biological significance test, which is about a comparison to the wild-type oscillation, distinct from the statistical amplitude test, which is exactly about significance with respect to noise. So, we begin by pooling genes that are similarly timed in the wild-type cell cycle, indicated by brackets in Fig. 2 C,D. For any given experiment, we consider when

the peak of all the genes in the same pool occurs by finding the time point, at which the majority of genes in the pool have their greatest peak. Thus, we filter out spurious peaks at random time points. In doing so, we follow the approach of (Pyne et al, BMC Genomics (2009)), where peaks that occur at different times cancel out. Then, we compute the ratio of the fold change of each gene at the pool's peak time point to the maximum fold change of that gene in the wild-type cycle. This ratio indicates the biological significance of the oscillation of any individual gene. (To assess the second peak for MET-CLN2-induced clnΔ^*clbΔ^* cells, we just pick the second peaks with no restriction that times have to match between different genes.) In each pool, almost all ratios lie on one side of the 10% or 30% thresholds and very few outliers on the other. One way to deal with the outliers would be to average over the ratios in each pool. We pursue a much more lenient strategy toward outliers. Outliers due to measurement or biological noise ought to be rare. So, we require that no less than 90% of the genes be on one side of the threshold or the other (see below for a probabilistic justification of the 90% rule). This suffices to allow a clear assignment of each of the pools to the categories significant/ambiguous/insignificant with the exception of the end-of-cell-cycle pool in the MET-CLN2 induced clnΔ^*clbΔ^* cells, where the difficulty in assigning one of these categories is a sign of interesting biology, we think, and which we explore in detail in the main text. Coincidentally, we find that a 90% rule similarly keeps the periodicity test, where we do not take the noise background explicitly into consideration either, from detecting too many false positives (see below), but we arrive at a 90% rule differently there.

To estimate the number of outliers we should expect above 10% of the wild-type oscillation just by chance, we find the number of genes in each pool, whose wild-type amplitude after reduction to 10% would have p values between 0.1-0.2, 0.2-0.3, or more than 0.3. (See below for p value calculations for amplitudes.) We multiply the numbers of these genes by 0.1, 0.2, and 0.3, respectively, and add them. We arrive at an underestimate for the number of genes, which would exceed 10% of the wild-type peaks just due to noise. For the three pools, the fractions of such false positives are 10.3%, 12.5%, and 10%, justifying our 90% rule.

A check on the choices we have made here is that, when applied to the data from [Orlando et al., 2008], the results are different from ours and in line with the conclusions reached by that article's authors.

Statistical amplitude test

The statistical amplitude test is only available for transcriptome data.

We slightly simplified the amplitude test of [de Lichtenberg et al., 2005], to make it more intuitive. Instead of comparing a fluctuation in a particular gene to fluctuations in all of the genes in the transcriptome, we compare the amplitude of an oscillation in a particular gene to fluctuations in the housekeeping genes, taken from Teste et al, BMC Molecular Biology (2009) and Cankorur-Cetinkaya et al, PLoS ONE (2012), which are not thought to be cell-cycle regulated according to [Spellman et al., 1998].

Specifically, these genes are: ACT1, ADH1, ALG9, ARF1, CCW12, CDC19, ENO1, FBA1, KRE11, GCN4, PDA1, PDC1, PGK1, RPS26A, TAF10, TDH3, TFC1, TPI1, UBC6.

Comparing to all other genes as in [de Lichtenberg et al., 2005] means that cell-cycle regulated genes are included in the basket of genes one is comparing against, which is problematic if the fraction of cell-cycle periodic genes is substantial. Furthermore, housekeeping genes have been extensively studied and serve as constitutively transcribed reference genes in the literature; housekeeping genes are not only non-cell-cycle regulated according to [Spellman et al., 1998], they are also considered to remain constitutively transcribed under changes in media, stress, etc.

We compile two cumulative distribution functions of fluctuations in housekeeping genes based on our data as well as for data from [Orlando et al., 2008]. Specifically, we compile a list of fold changes $(x_i/x_1 - 1)$ for each housekeeping gene, where x_i is the FPKM score (our data) or the publicly available microarray scores (Orlando et al.) at time point i where $i > 1$. The means (-0.04,-0.0009), standard

deviations (0.235,0.187), and p=0.01 thresholds are similar (0.56,0.62) for the two distributions, indicating that background fluctuations are comparable. Given that oscillations with low p values are particularly important, our data set is slightly less noisy by that standard since a $p < 0.01$ oscillation only has to be greater than +56% compared to +62% in the [Orlando et al., 2008] data.

To assign a p value to the oscillations in the transcriptome data, we proceed at first as for the biological significance test for each pool, determining the peak time for the pool by finding the time point at which the majority of genes have their greatest peak. (Again, for the second peak, there is no restriction that the times have to be the same within the same pool.) Then, unlike the biological significance test, we compute the p values for each of the peaks in the same pool and combine the p values by Fisher's method. The threshold for significance is p=0.01.

Again, as we discuss in the main text, when applied to the data from [Orlando et al., 2008], the results are different from ours and in line with the conclusions reached by that article's authors.

Statistical periodicity test

We apply the test exactly as described in [de Lichtenberg et al., 2005]. For any individual time course, random permutations of the data are generated, and for each permutation, the fit to a sine and cosine function is calculated. The p value is the fraction of permutations that have a better fit than the original time course. A specific period for the sine and cosine functions is required as a fixed parameter; we cover all periods between 50' and 130' at intervals of 5' (including 50' and 130') to cover a large range.

For the transcriptome data we perform 10^6 random shufflings for each gene unless there are only n time points such that n! is smaller than 10^6, in which case we consider all n! permutations. We obtain a p value for each gene in each experiment. We do not apply the periodicity test to the transcriptome measurements on Cln-blocked cells (Fig. 2 C) because there are too few time points for such a test to make sense; oscillations have to be ruled out or confirmed by the other two tests. (We do investigate periodicity in Cln-blocked and Cln,Clb-blocked cells at high temporal resolution at the single-cell level instead (Fig. 3 B,C).)

For the single cell data, we perform only 10^5 random permutations of each cell's time course because there are many single cell time courses available for the same gene, which achieves high accuracy through pooling. In any case, in individual case studies, we find almost no relevant difference in p values using 10^5 or 10^6 random permutations. We use the raw, unsmoothed data over the same time span shown in the figures to obtain a p value for each single-cell/single-gene time course.

The periodicity test, as described in [de Lichtenberg et al., 2005], is very sensitive; it describes how well a given time course matches a sine or cosine function compared to random permutations of the same time course. In some cases, this may be an unreasonably low standard for periodicity. For a few single-cell time course experiments, we suspected that the test is detecting features that have nothing to do with periodicity. For example, in Fig. 3 H,I, even heavily smoothed time courses are detected as significantly periodic, even though they clearly have no repeat 'bumps' after smoothing. We found that there are two scenarios, in which we have to either modify the test or substitute a similar test. In some cases, the first (expected) oscillation suffices to yield a low p value for periodicity, e.g., in Fig. 3 H. In these cases, we compute the periodicity test for the time course after the expected, normal oscillation, i.e., roughly after 100' in Fig. 3 H. In other cases, the underlying trend alone gives a low p value. For example, the slow increase and decrease after 0' in Fig. 3 I has a low p value for periodicity even after heavy smoothing removes all fluctuations around the slowly changing trendline. In these cases, we fit a smooth line, which shows no oscillations, through the time course. We compute the mean and standard deviation of the difference between the actual time course and the smooth approximation. We then generate 10^5 time courses with Gaussian random noise on top of the smooth line with the same mean and standard deviation as the actual time course. For each, we compute the fit to a sine or cosine function just like

in the periodicity test of [de Lichtenberg et al., 2005] and assign p values in the same way; again, the p value is the fraction of random time courses that fit a cosine or sine function better than the original time course. These two modifications to the periodicity test were only needed in Fig. 3 H,I, wild-type cell cycles in Fig. 4 A,B, and swi5Δ cells in Fig. 4 G-I.

Before we can combine p values of each gene within a pool of genes for transcriptome data or the p values of single cells for the same gene for single-cell/single-gene data using Fisher's method, we need to correct for spurious false positives. We have not discounted fluctuations which are small, as is commonly done by combining the periodicity test with the amplitude test [de Lichtenberg et al., 2005] (Marguerat et al, Yeast (2006); Pyne et al, BMC Genomics (2009)). In contrast to those articles' authors, we would like to keep the periodicity and the amplitude tests distinct. (We faced the same challenge with the biological significance test, see above.) Yet, the problem of too many false positives has been noted in the past even when the amplitude test has been included. Different approaches have been taken to deal with this problem by those articles' authors.

To inform our approach, we performed the periodicity test on the housekeeping gene data. In one replicate of the wild-type cell cycle in [Orlando et al., 2008], the collection of housekeeping genes turns out to be significantly periodic. Neither the other wild-type replicate, nor the two clb1-6Δ GAL1-CLB1 replicates from [Orlando et al., 2008], nor our two MET-CLN2-induced cell populations show significant periodicity in the collection of housekeeping genes for any period. We conclude that this is a false positive, as would be expected for housekeeping genes. We decided to remedy the high false positive rate by accepting that there can be outliers with exceptionally low p values lowering the cumulative p value of the entire pool. We found that by removing at least 10% (2 out of 19 genes) of the lowest individual p values, the false positive p value for the housekeeping genes in the Orlando et al. data would be fixed.

Thus, we apply Fisher's method to combine individual p scores, but we remove the lowest 10%, similar to the biological significance test. So, statistical significance must be supported by any 90% of the data. Then, we consider periodicity statistically significant if it is significant ($p< 0.01$) for any period (50'-130') that we test.

As with the other tests, when applied to the data from [Orlando et al., 2008], the results are different from ours and in line with the conclusions reached by the article's authors. Specifically, periodicity in Start-G2 genes is detected as claimed.

Image analysis

Automated image segmentation and fluorescence quantification of yeast grown under time-lapse conditions were performed as previously described. [Charvin et al., 2010] Images were taken every 5min in all experiments except in no-Clb experiments, where images were taken every 10min.

Scatter plots

If a cell's ith peak does not occur before the end of the movie, $\Lambda_{\uparrow,\downarrow i}$ is set to zero. In some cyclin-depletion experiments (where the timelapse recording is long ($\approx$ 10 h)), the maximum waiting time for the first peak was reduced to $300'$ to ignore fluorescence fluctuations that are very late and thus not relevant for cell cycle oscillations.

Data processing - Smoothing

Time courses were smoothed using Matlab's SmoothingSpline function. Based on the strength of the fluorescent signal, noise, and the image acquisition intervals, we adjusted the smoothing parameter but used the same parameter consistently in related experiments: $2 * 10^{-5}$ for all no-Clb experiments except

with SWI5-YFP ($5 * 10^{-5}$, less smoothing). $2 * 10^{-4}$ for recordings of cycling cells and for all high-Clb experiments except SWI5-YFP in the cdc20Δ block ($1 * 10^{-3}$, less smoothing). $4 * 10^{-4}$ for SIC1pr-YFP recordings in GAL1-SIC1 experiments. $5 * 10^{-4}$ for Clb2-YFP recordings in clnΔ^* GAL1-SIC1 experiments. An example of a raw and smoothed time course for the same cell is shown in Fig. 1 A.

Modeling

For Fig. 5, the model by Chen et al. [Chen et al., 2004] was obtained from the website of the Tyson lab. The simulation was run and interrupted at the first minimum after the second peak of Clb2. The term proportional to [mass] in the accumulation rate of Clb2 was then multiplied by various constants smaller than than 1. Smaller constants lead to smaller and later Clb2 peaks.

A pulse (orange) was added to Swi5 in the equation governing the evolution of Sic1 at the time points indicated in the plots. The area of the pulses was chosen to be 37conc./min, which is equal to the area (integral of height over time) of SWI5 over one cell cycle in that model. Each pulse was modeled by an inverse parabola with exponent 16.

Western blotting and kinase activity assay for Clb2

Log-phase liquid cultures were grown in G-Met overnight, then switched to the media indicated in the figures and the text. Parallel cultures were kept in G-Met for the duration of the time course. The equivalent of 5 mL at 0.3 OD660 was collected from each culture and cells were washed once in LSHN (50mM NaCl, 10mM HEPES pH 7.5, 10% Glycerol). Glass beads (400uL) and 400uL LSHNN (LSHN + 0.1% NP-40) with proteinase and phosphatase inhibitors were added, and cells were broken with a FastPrep FP120 Cell Disruptor (Thermo Electron Corp.) in two 20s intervals at setting 5. Cell lysates were centrifuged for 2min, and 10uL of the total lysate was transferred to 50uL sodium dodecyl sulfate (SDS) sample buffer with beta-mercaptoethanol. Western blotting was carried out with these total lysates by use of standard methods. Anti-Clb2 (Covance) and anti-Pgk1 (Invitrogen) were both used at 1:10,000 dilution. HRP-conjugated secondary antibodies (GE) were used at 1:4,000 dilution. Blots were imaged on a Fujifilm LAS-3000 imager. The remaining cell lysates were immunoprecipitated with anti-Clb2 antibody (Covance) at 1:700 dilution. Kinase activity assay was carried out with Clb2 immunoprecipitates essentially as described in (Levine, Huang, Cross, Mol. Cell. Biol. (1996)). Histone H1 radioactivity was detected using a Typhoon 9400 variable imager (Amersham Biosciences). Both Western blot and kinase activity images were quantitated using Image J software (Schneider, Rasband, Eliceiri, Nat. Methods (2012)).

DNA content measurements

Standard methods were used for measurement of DNA content by flow cytometry. Cells were fixed in 70% ethanol, washed (50 mM Tris, pH 7.8), treated with RNAseA (2 mg/ml), then treated with pepsin (5 mg/ml in 55 mM HCl), and, after two more washes, stained with propidium iodide (1ug/ml). DNA content was measured on a BD Accuri C6 flow cytometer.

Scoring for Fig. 6

We only consider cells that after the switch to glucose had one nucleus and no budneck marker before budding again. Of the cells we scored, those that rebudded before nuclear division, which occurred occasionally with clb2Δ and clb2Δsic1Δ cells, were counted as failing to complete the cell cycle (time to nuclear division or rebudding infinite).